

I SEMINÁRIO DE SUPORTE À PESQUISA E GESTÃO  
DE DADOS CIENTÍFICOS

*Panorama Atual e Desafios*

Universidade Federal de Santa Catarina (UFSC)



# eScience: desafios e possibilidades

Renata Gonçalves Curty

Departamento de Ciência da Informação/UFL

# eScience

(Digitally) *Enhanced Science*

Termo originalmente atribuído ao inglês **John Taylor**, quando em **1999**, na posição de diretor do escritório de **Ciência e Tecnologia do Reino Unido**, fez o anúncio de um programa nacional multimilionário que iniciaria em 2000, o qual tinha como objetivos ampliar a **cooperação global** em importantes áreas do conhecimento e estabelecer a nova geração de **infraestrutura computacional**.

# IEEE International Conference on eScience

[Home](#) [Steering Committee](#)



- eScience Workshop 2010
- eScience Workshop 2011
- eScience Workshop 2011-Transforming Scholarly Communication
- eScience Workshop 2012
- eScience Workshop 2013
- eScience Workshop 2014

## *Papel pervasivo da tecnologia na produção de ciência e o protagonismo dos dados científicos*



[Workshop Webcats](#) [Speaker Biographies](#)

### Workshop

Microsoft Research hosted a two-day e-science workshop on Thursday, 2005 and Friday, October 7, 2005 in Redmond, Washington. This workshop was a follow-on workshop to the successful SciData 2004

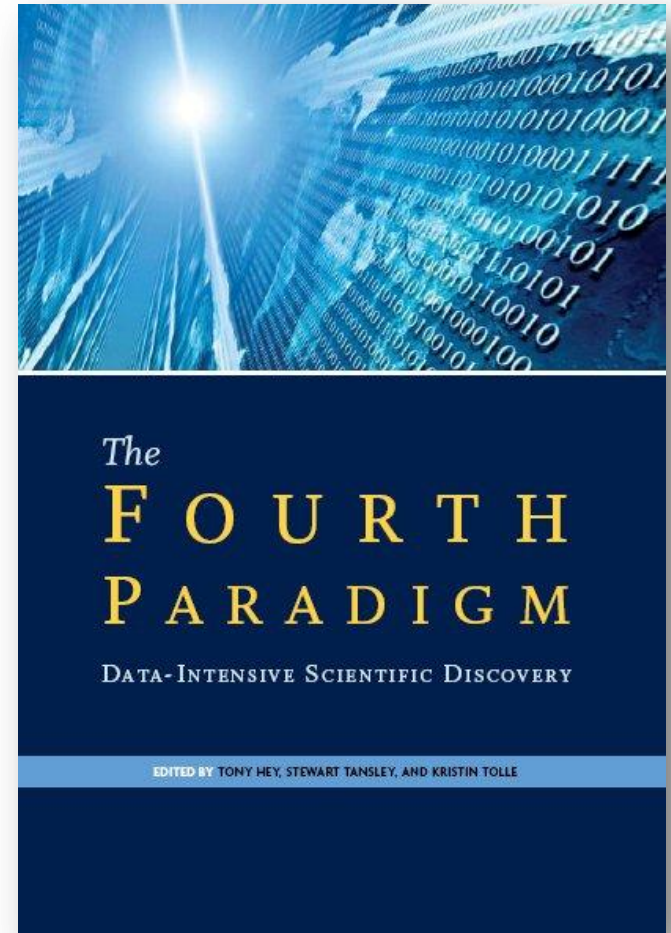
The eScience Workshop provided a unique opportunity to learn and affect what is happening in the realm of data intensive scientific computing within Microsoft. Attendees learned first hand from early adopters using Microsoft Windows, Microsoft .NET, Microsoft SQL Server, and Web services in these problem spaces, as well as explored in-depth how modern database technologies and techniques are being applied to scientific computing. By providing a forum for scientists and researchers to share their experiences and expertise with the wider academic and research communities, this workshop fostered collaboration, facilitated the sharing of software components and techniques, and established a vision for Microsoft Windows and .NET in data intensive scientific computing.



# eScience >> Data Science

## Ciência Orientada a Dados

- Intensivamente computacional (*Grid Computing*)
- Dados abertos e disponíveis online
- Colaboração via redes distribuídas
- Maior transdisciplinaridade
- Maior transparência, dinamismo, e produtividade



*eResearch?*  
*Data Research?*

# CONJUNTURA

## Ciência Aberta



## Políticas & Mandados Institucionais



## Plano de Gerenciamento de Dados



# COMPARTILHAMENTO DE DADOS

Pesquisa com 1.317 cientistas de 43 disciplinas nos EUA

## Relação Positiva

- Pressão regulatória por periódicos
- Pressão normativa (disciplina)
- Benefício de carreira percebido
- Altruísmo

## Relação negativa

- Esforço percebido

## Sem significância

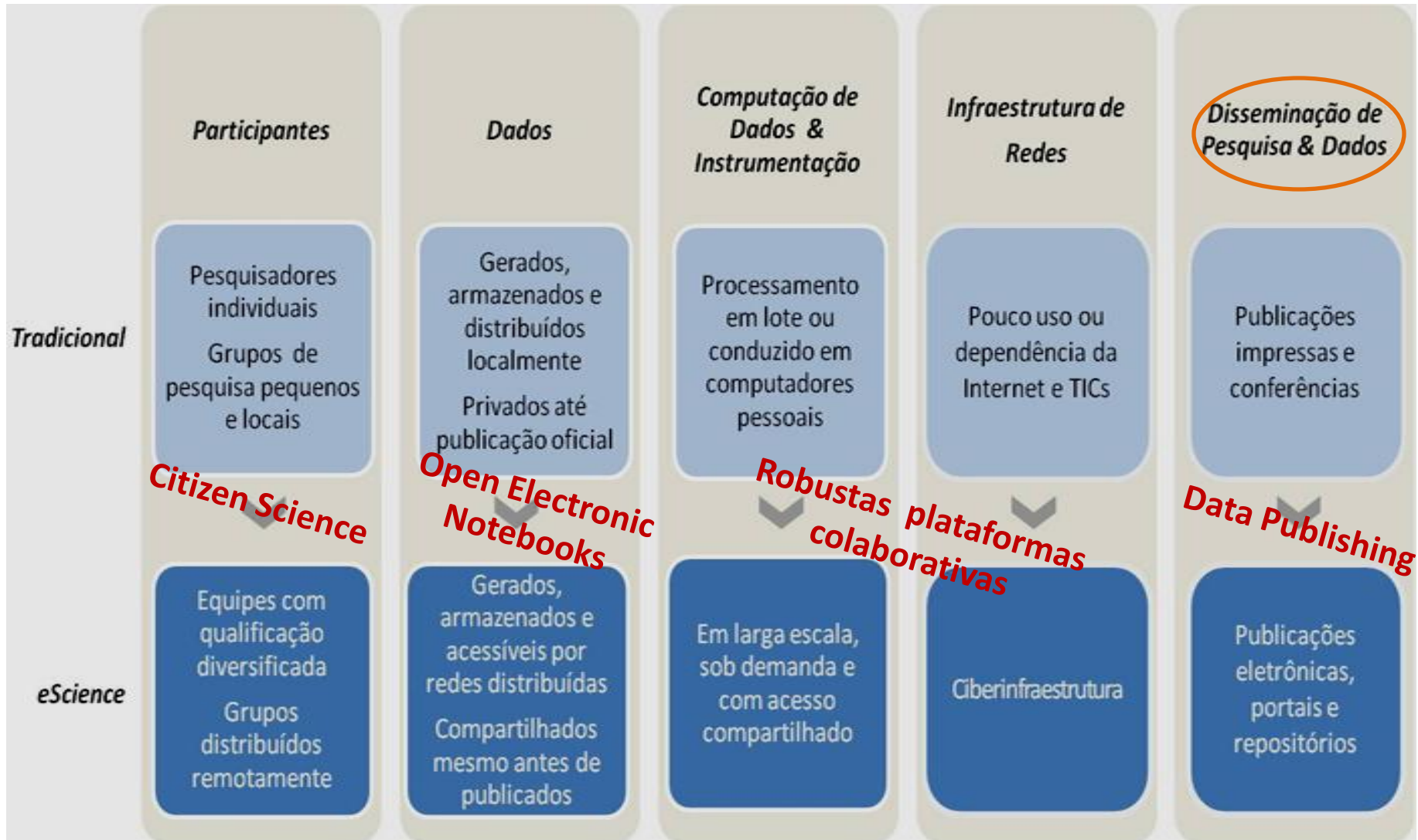
- Pressão por parte das agências de financiamento
- Disponibilidade de repositórios de dados



Toda nova aplicação de um conjunto de dados ou de diferentes conjuntos de dados de forma combinada, com objetivo de **verificação**, **replicação**, ou para extração de novos *insights*, **testes de novos pressupostos ou hipóteses** a partir de novas perguntas de pesquisa, podendo ser realizado por terceiros ou por pesquisadores envolvidos na coleta de dados do estudo original.



# Ciência vs. eScience



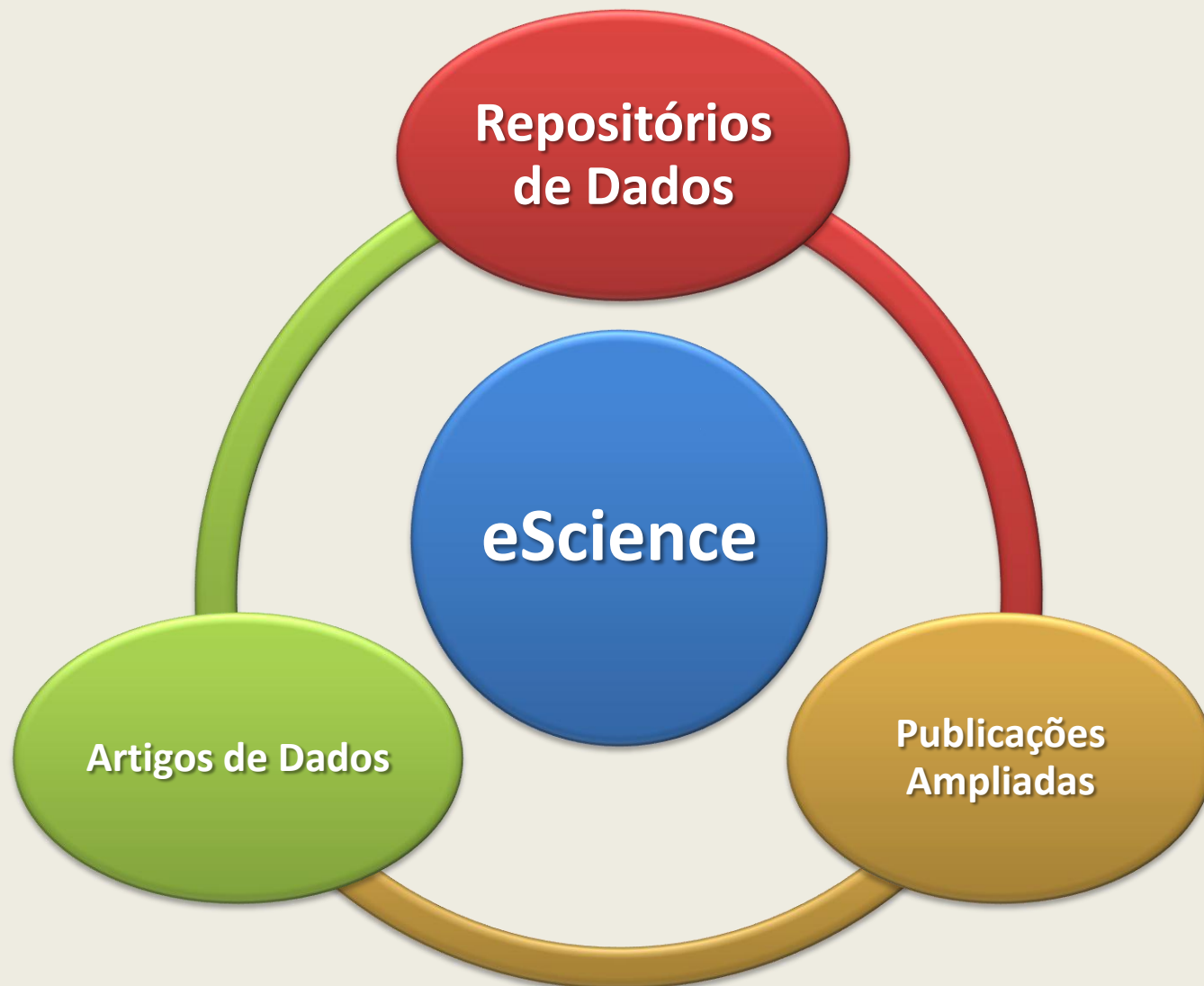
CURTY, R. G. eScience: diferentes vieses, fontes e iniciativas. In: TOMAEL, M. I.; ALCARÁ, A. (Org.). *Fontes de Informação Digital*. Londrina: EDUEL, 2016. p. 77-118.

APPELBE, B.; BANNON, D. eResearch: paradigm shift or propaganda? *Journal of Research and Practice in Information Technology*, Sidney, v. 39, n. 2, p. 83-90, 2007.



PUBLICATIONS AND DATA

# FORMAS E FORMATOS PARA DISSEMINAÇÃO DE DADOS CIENTÍFICOS



# PUBLICAÇÕES AMPLIADAS



**Disseminação  
Publicações  
Convencionais**  
(visível aos pares)

**Processo de Pesquisa**  
Concepção até  
disseminação  
(invisível aos pares)

# PUBLICAÇÕES AMPLIADAS

Buscam **ampliar o potencial das produções científicas** por meio de **dados interligados** (*linked data*) e minimizar as dificuldades de pesquisadores que buscam maior detalhes sobre o processo da pesquisa, bem como acesso aos dados analisados.



# DIFERENTES NÍVEIS

1. Interligação de **ativos externos** e de **estudos relacionados**
2. Capacidades de interação por meio de **recursos Web 2.0** pós-publicação (comentários, recomendação, ranking, avaliação)
3. Interligação de **dados da própria pesquisa** (evidência) com objetivo de ilustrar ou clarificar informações
4. Possibilidade de reprodução com **dados executáveis** e **recursos de interação** para plotagem e visualização de dados

# PUBLICAÇÕES AMPLIADAS

## Vantagens

- ✓ Redução na duplicação de pesquisas
- ✓ Melhor contextualização dos dados pela própria publicação, mantendo seu sentido original
- ✓ Maior transparência e possibilidade de verificação dos dados e resultados/análise no momento da leitura (revisão pelos pares e leitores)
- ✓ Potencializa a descoberta de pesquisas relacionadas
- ✓ Amplia a possibilidade de colaboração em pesquisas
- ✓ Maior possibilidade de impacto e citação





# EXEMPLOS

This screenshot shows the ScienceDirect article page for "Importance of Ethiopian shade coffee farms for forest bird conservation". The article is by Evan R. Beachley, Çağan H. Şekeroğlu, Anagaw Abdani, Getay Gebremichael, James Kuna Ndung'u, Birtukan Abdu Mahamud, Tassew Beyene, Tarku Mekonnen, and Luc Lent. The abstract states that coffee is a major tropical commodity and that shade coffee farms in Ethiopia have a higher species richness of forest birds compared to open coffee farms. The article is part of a Special Issue Article titled "Tropical Insectores".

<http://www.sciencedirect.com/science/article/pii/S006320715000324>

This screenshot shows the ScienceDirect article page for "Estimating the costs of medicalization" in the journal "Social Science & Medicine". The article is by Peter Conrad, Thomas Mackie, and Abey Mahotra. The abstract discusses how medicalization is the process by which non-medical problems become defined and treated as medical problems, leading to increased health care costs. The article is part of a Special Issue titled "Medicalization in the 21st century: Introduction".

<http://www.sciencedirect.com/science/article/pii/S0272953610001784>

This screenshot shows the ScienceDirect article page for "Possible misleading interpretations on magnetic and transport properties in BiFeO<sub>3</sub> nanoparticles caused by impurity phase". The article is by Fabian E.N. Ramirez, Gabriel A.C. Pasca, Jose A. Souza, and others. The abstract discusses the synthesis of BiFeO<sub>3</sub> nanoparticles and the potential for misleading interpretations of magnetic and transport properties due to impurity phases. The article is part of a Special Issue titled "Physics Letters A".

<http://www.sciencedirect.com/science/article/pii/S0375960115003102>

This screenshot shows the PhytoKeys article page for "A nonet of novel species of *Monanthotaxis* (Annonaceae) from around Africa". The article is by Paul H. Hoekstra, Jan J. Wieringa, and Lars W. Chatrou. The abstract describes nine new species of *Monanthotaxis* from West Africa, Central Africa, and Tanzania. The article includes a map of Africa showing the distribution of the new species. The article is part of a Special Issue titled "Research Article".

<https://phytokeys.pensoft.net/article/9292/list/6/>



# DESAFIOS

- ➡ Identificadores persistentes globais únicos para resolução das interligações
- ➡ Interoperabilidade e formatos (reuso futuro)
- ➡ Indicações de citação e mecanismos para proteção e atribuição dos dados
- ➡ “Propriedade” dos dados e a preocupação com a detenção deles por grandes editores científicos (modelos de negócios)



**E no Brasil?**

# DATA PAPER

- Documento de metadados pesquisáveis, descrevendo um conjunto de dados em particular ou um grupo de conjuntos de dados, publicados na forma de um artigo *peer-reviewed*.
- Ao contrário de um artigo de investigação analítico convencional, o objetivo principal de um artigo de dados é o de descrever os dados e as circunstâncias de sua coleção, ao invés de apresentar hipóteses e conclusões.



# EXEMPLO

<https://phytokeys.pensoft.net/articles.php?id=1551>

**PhytoKeys** Articles About About Pensoft Books E-Books Blog Journals Register | Log

Data Paper PhytoKeys 25: 55-67 (24 Jul 2013)  
<https://doi.org/10.3897/phytokeys.25.3100>

## Database of Vascular Plants of Canada (VASCAN): a community contributed taxonomic checklist of all vascular plants of Canada, Saint Pierre and Miquelon, and Greenland

Peter Desmet<sup>1</sup>, Luc Brouillet<sup>1</sup>

<sup>1</sup> Université de Montréal Biodiversité Centre, 4101 rue Sherbrooke est, H1X2B2, Montréal, Canada  
Corresponding author: Luc Brouillet ([luc.brouillet@umontreal.ca](mailto:luc.brouillet@umontreal.ca))  
Academic editor: Vishwas Chavan  
Received 19 March 2012 | Accepted 17 July 2013 | Published 24 July 2013

(C) 2013 Peter Desmet. This is an open access article distributed under the terms of the [Creative Commons Attribution License 3.0 \(CC-BY\)](https://creativecommons.org/licenses/by/3.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.  
For reference, use of the paginated PDF or printed version of this article is recommended.

**Resource citation:** Brouillet L, Desmet P, Coursol F, Meades SJ, Favreau M, Anions M, Bélisle P, Gendreau C, Shorthouse D and contributors\* (2010+). Database of Vascular Plants of Canada (VASCAN). 27189 records. Online at <http://data.canadensys.net/vascan> <https://doi.org/10.5886/175M7V5P> and <http://www.gbif.org/dataset/3f8a1297-3259-4700-91fc-acc4170b27ce>, released on 2010-12-10, version 24 (last updated on 2013-07-22). GBIF key: [3f8a1297-3259-4700-91fc-acc4170b27ce](https://doi.org/10.3897/phytokeys.25.3100). Data paper ID: doi: [10.3897/phytokeys.25.3100](https://doi.org/10.3897/phytokeys.25.3100)

\* see Appendix.

The article is also available on GitHub at <https://github.com/peterdesmet/vascan-data-paper>

Contents Article info Citation Metrics Comment Related  
Figs Tabs Refs Cited

- Top
- Abstract
- Data published through
- Project details
- Taxonomic coverage
- Spatial coverage
- Temporal coverage
- Sampling methods
- Dataset
- External datasets
- Acknowledgements
- References
- Appendix

Home About Contact Content Research Integrity Search... Log in Register

## Journal of open health data

Start Submission Become a Reviewer

Reading: The European Distribution of *Sus Scrofa*. Model Outputs from the Project Described within the... Share: f t g in

Download A- A+ 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

### Data papers

## The European Distribution of *Sus Scrofa*. Model Outputs from the Project Described within the Poster – Where are All the Boars? An Attempt to Gain a Continental Perspective

**Authors:** Neil Stewart Alexander ✉, Giovanna Massei, William Wint

### Abstract

Wild boar is a host of a number of arthropod-vectored diseases and its numbers are on the rise in mainland Europe. The species potentially impacts ecosystems, humans and farming practices and so its distribution is of interest to policy makers in a number of fields beyond that of the primarily epidemiological goal of this study.

**JUMP TO**

- Abstract
- 1. Overview
- 2. Context
- 3. Methods
- 4. Dataset description
- 5. Reuse potential
- Competing Interests
- Acknowledgements
- References

<http://openhealthdata.metajnl.com/articles/10.5334/ohd.24/>

# DATA PAPER - PROPÓSITOS

- Descrever os **dados de forma estruturada, legível** por máquinas e também **inteligível** por humanos
- **Avaliação de dados pelos pares** (credibilidade)
- Compor uma publicação citável de modo a oferecer **crédito aos autores e aos produtores dos dados**
- **Promover** e trazer à atenção da comunidade acadêmica a existência de **dados de pesquisa com potencial de reuso**
- Monitorar de forma mais eficaz o uso e **citações dos dados de pesquisa**

# RESSALVA...

- Artigos de dados são adequados para descrever dados relativamente estáveis e em menor escala
- Dados em larga escala, coletados em tempo real e em fluxo contínuo (pesquisas climáticas, marinhas e astronômicas) esta modalidade de publicação de dados torna-se menos apropriada.



# DESAFIOS

- 👉 Tornar o processo de publicação de artigos de dados **mais fácil e ágil**
- 👉 Definir **estratégias de visibilidade e espaço** para os papers de dados
- 👉 Discutir a relação de **crédito e recompensa** para esse tipo de publicação (descritivo vs. analítico)



# REPOSITÓRIOS DE DADOS



Cerca de **1900 repositórios**, vinculados à instituições legais de pesquisa, provenientes de diferentes países e nas mais variadas áreas do conhecimento, dedicados ao compartilhamento de dados.

Três repositórios mantidos pelo governo brasileiro:

- **Banco de Dados de Exploração e Produção (BDEP)**, na grande área de Ciências Naturais;
- **Repositório de Dados de Levantamentos Biológicos (Ppbio)**;
- **DataVerse Network** gerenciado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).





**Se construirmos, eles virão...**

Mais do que  
ciberinfraestrutura...





## COMMUNITY-BASED

Inexistência de políticas norteadoras  
Deslizes no processo de implantação  
Falta de estratégias de sustentabilidade e  
engajamento

Foco nas duas pontas  
(quem compartilha, quem reusa)



# REPOSITÓRIOS DE DADOS

**Necessidade de conciliar o melhor dos “dois mundos” ...**

- Facilidade de uso/simplicidade
- Recompensa (Badges)
- Visibilidade
- Métricas Alternativas



The **Dataverse** Project 

- Validação e endosso (metadados, dados e documentação suplementar)
- Publicações relacionadas e derivações
- Citação e atribuição
- Controle de versões e proveniência

Identificadores únicos  
ORCID, DOI, Funder ID, Grant ID...

# BADGES (DISTINTIVOS)



Open Data



Open Materials



Simple intervenção que estimula o compartilhamento de dados e materiais, bem como maior engajamento dos cientistas em plataformas colaborativas.

# DESAFIOS

- Entendimento aprofundado das **comunidades científicas e suas especificidades** (tipos e natureza dos dados produzidos, valores e normas, etc.)
- Aspectos **éticos e legais**
- Estratégias para promoção do **auto-arquivamento e compartilhamento**
- Mecanismos de **incentivo** e sistemas de **recompensa**

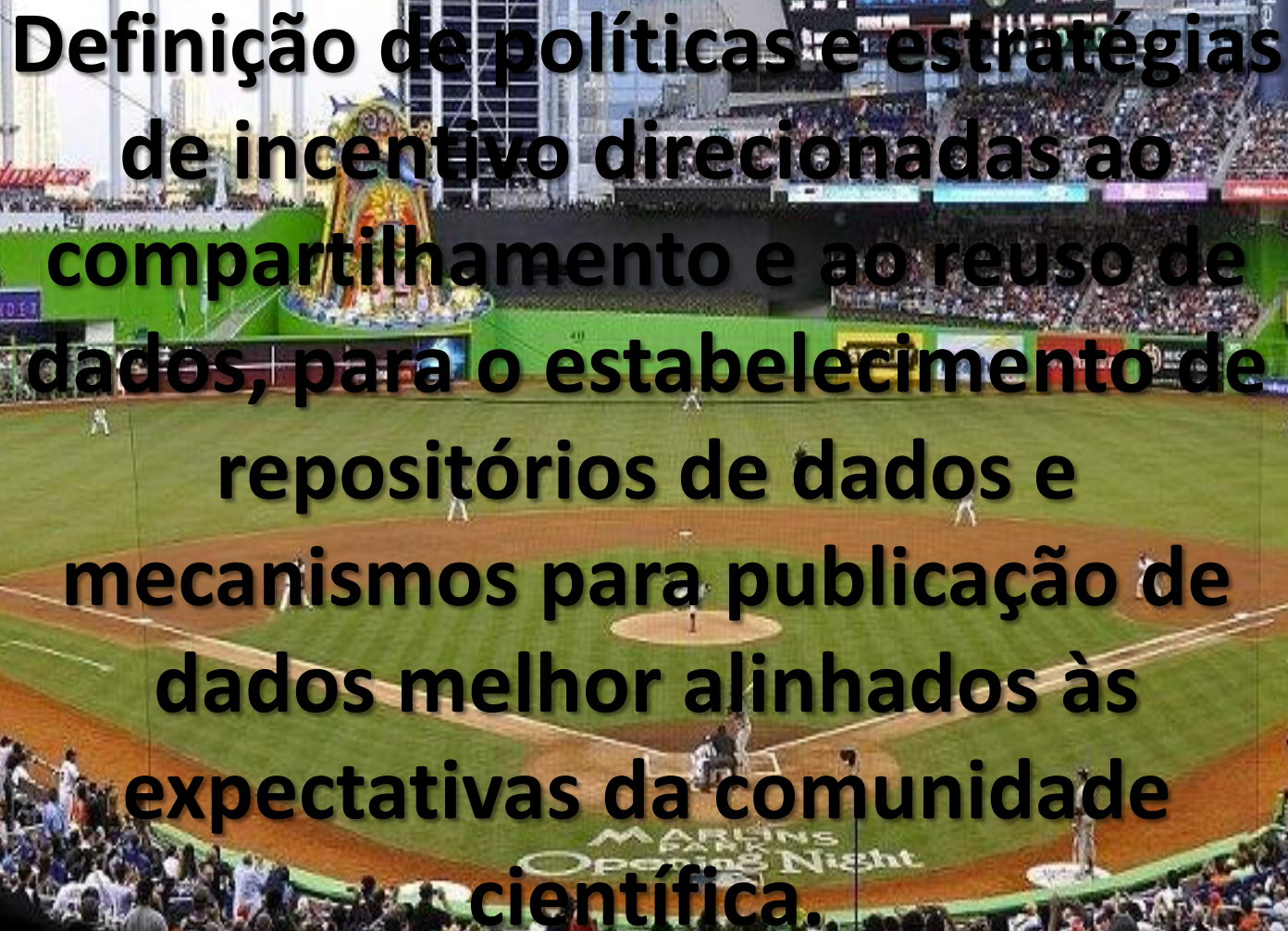


# DESAFIOS

- Estratégias para **curadoria e preservação** de dados (longo prazo)
- Otimização da **encontrabilidade** de dados
- Documentação e metadados que possibilitem a devida **contextualização para efetivo reuso de dados**







**Definição de políticas e estratégias de incentivo direcionadas ao compartilhamento e ao reuso de dados, para o estabelecimento de repositórios de dados e mecanismos para publicação de dados melhor alinhados às expectativas da comunidade científica.**



# **GESTÃO DE DADOS CIENTÍFICOS**

## **NOVOS RUMOS**

**MAIOR AUTOMAÇÃO DOS PROCESSOS**  
**REDUÇÃO DE ESFORÇOS PARA O COMPARTILHAMENTO DE DADOS**  
**MENOR FRICÇÃO NO PROCESSO DE REUSO**





# **Machine-actionable Data Management Plans (maDMPs)**

## **Automated Data Papers**

**Obrigada!**

recurty@gmail.com